The ESO Data Management Division

P. QUINN, ESO Data Management Division

Astronomers are currently at an important cross-road in their ability to gather information about the Universe. Firstly, computers and astronomical detectors have simultaneously increased in performance. Modern observatories using high-throughput, large-format detectors for imaging and spectroscopy are capable of producing tens of gigabytes of raw data every clear observing night. Although this allows rapid progress on observing programmes, the data rate and volume can easily saturate current network technologies and consume the disk space typically found on most workstations. Secondly, large telescopes in the 8–10-m class are going to be the workhorses of optical and IR astronomy well into the next century. These telescopes will be available to astronomers around the globe and competition for these expensive resources will be fierce. Observing programmes then have to be undertaken in the most efficient manner possible, minimising the effects of the weather in order to complete programmes in the shortest possible time.

Because of this "data deluge" and the need for maximum efficiency, astronomers are faced with the realisation that perhaps the classical observing cycle of trekking to the telescope and returning home with a tape of unprocessed data is no longer the most efficient or practical way to do astronomy. ESO astronomers stood at this cross-road and considered the options carefully when designing the Science Operation Plan for the VLT.

In this plan, ESO decided to enable the VLT to execute observing programmes in a service mode. Astronomers can specify programmes in a Phase I–Phase II proposal process and the resulting observations would be completed by ESO astronomers and technicians at the VLT when the conditions are most favourable for that programme. In this way, ESO astronomers would be as efficient and competitive as possible in completing the types of frontier programmes that will be common place on the VLT. Furthermore, ESO decided to meet the challenge of the data volume potential of the VLT by insisting that the VLT should be able to deliver calibrated data to the astronomer with a well-defined accuracy. This has two immediate benefits for the astronomer. Firstly the data processing and storage necessary to calibrate the data is provided by ESO which helps relieve the "deluge" significantly since most data reduction systems expand raw data by factors of 2–5 before calibration is complete. Secondly, by guaranteeing the calibration process to a given accuracy, ESO has to continuously monitor the applicability of calibration data and the performance of telescopes and instruments over long periods of time. The quality control of the calibration process will ensure the long-term usefulness of the data.

Once ESO adopted the principles of service observations and delivery of calibrated data for the VLT, it was obvious that the flow of data must be managed from start to end of the observing process. The ESO Data Management Division was formed to design and implement the Science Data Flow System for the VLT. Beginning in 1994, the division was formed from pre-existing groups at ESO whose responsibilities covered proposal submission, data archiving and the processing of ESO telescope data using the MIDAS system.

By the end of 1995, the division consisted of three main components: the Science Archive Group, the Data Pipeline Group and the User Support Group. Each of these groups addresses specific components in the Data Flow System (DFS).

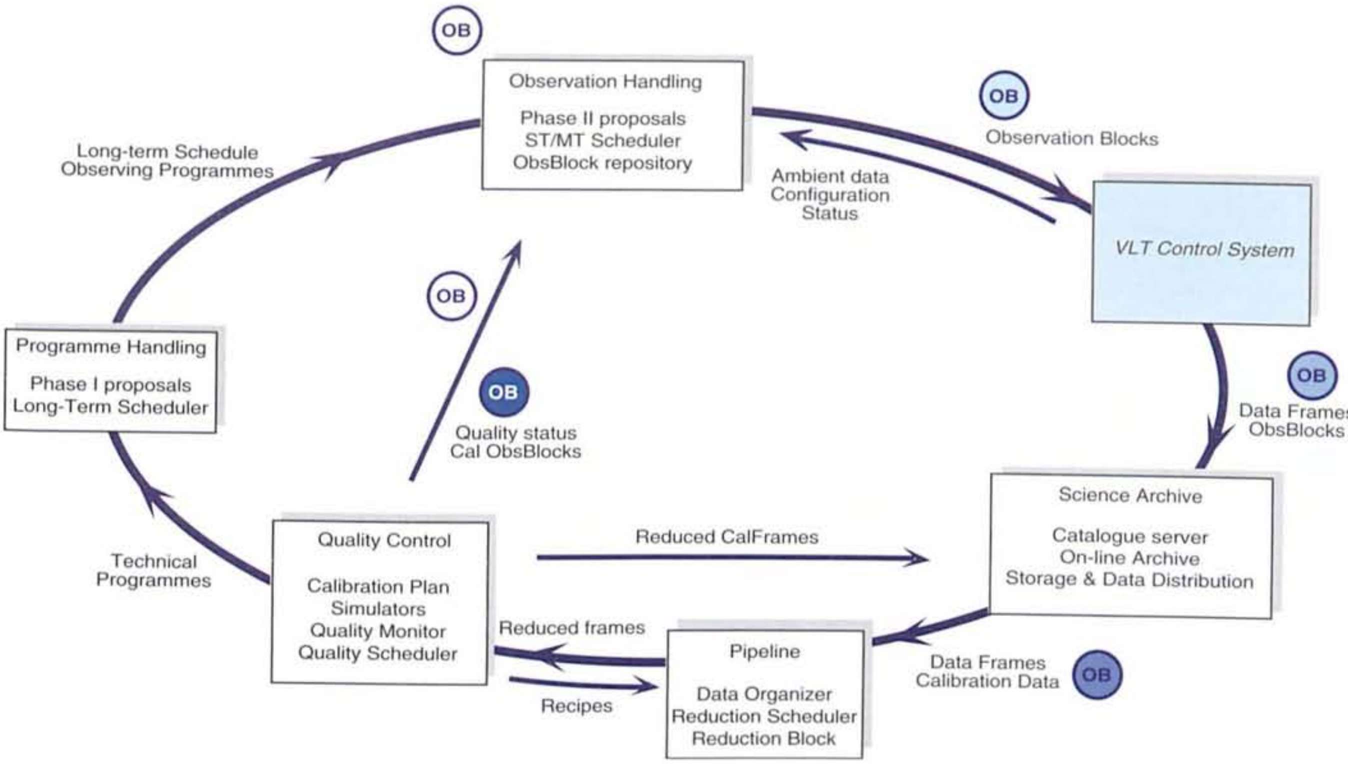
The Design of the DFS

In early 1995, the Data Flow Working Group was formed to formulate the requirements on the DFS following the guidelines of the Science Operation Plan. The group consisted of astronomers, instrumentation and software experts from several ESO divisions under the leadership of Preben Grosbøl. Discussions within the group focused on the characterisation of the "observing cycle" from proposal entry to final archive of the data. This cycle was broken down into component processes and the outline of the DFS took shape. A fundamental concept that holds the component processes together is that of the Observation Block. This "quantum of data" in the data flow system is an object which contains the essential target and instrument information to define an observation of a single object. An Observation Block is defined by either a classical observer at the telescope or a service mode observer. Whether the observer uses service-mode capabilities or not, the Observation Block is intrinsic to the flow of data from place to place in the VLT operation. Preliminary design documents describing the services offered by each of the modules in the DFS and their interrelationships were drawn up and distributed by the middle of 1995. The basic modules in the DFS are:

(1) Programme Handling:	The electronic submission of Phase 1 proposals. Preparation of a 6-month schedule for service and classical mode observing sessions.
(2) Observation Handling:	Phase II proposals are prepared using software tools. Observation Blocks are created for each target. Medium-Term Scheduler system prepares a list of possible Observing Blocks to be executed for the next few nights. Short-Term Scheduler systems select the next Observation Block to be executed by the VLT Control System (VCS) based on environment and configuration data.
(3) Science Archive:	Provides on-line catalogue services to observers. Stores raw data passed by the VCS. Data distribution and archival research on VLT data.
(4) Pipeline:	Organise raw data and calibration information for analysis. Schedule calibration processing of data. Apply approved calibration recipes.
(5) Quality Control:	Maintain calibration plan for a given instrument and mode. Compare calibration data with simulators. Study long-term trends in calibration data. Assess quality of calibrated data with respect to standards.

ESO VLT Data Flow Subsystem Breakdown

Data Management Division, April 1996



The DFS is the scientific operating system of the VLT. The VLT can be viewed as one machine with multiple components that interact. Unless the components of the VLT machine work together in an efficient manner, the scientific potential of the VLT will not be fully realised. With the DFS as a science operating system, the VLT machine can be scheduled and can make use of redundancy, pipeline and aggregation of resources to maximise throughput. In the design of the DFS, many opportunities exist to exploit the single machine vision of the VLT.

The concept of the DFS arrived late in the process of designing and implementing of the VCS. The VCS system is responsible for the control of the VLT Unit Telescopes and their associated instruments. This responsibility extends from low-level motor control to high-level assessments of the technical quality of data and its movement from instruments to storage. VCS was designed and partially implemented before concepts like Observation Blocks were introduced. It is therefore clear that the success of the DFS will depend critically on the close collaboration of software and system designers from DMD and the other ESO divisions. In September 1995, a Data Flow Project Team was formed by the Data Management Division to bring together astronomers and engineers from the VLT, Science and Instrumentation Division to address the many interfaces issues opened between the DFS and VCS. This Team has worked hard to crystallise the critical system interface issues and a number of collaborative software programmes are now under way between DMD and the VLT software group. A recent review of the status of VCS and DFS design in April 1996 (see this volume) pointed out the importance of the ongoing work of the Data Flow Project Team.

The development of a close collaboration and coupling between VCS and DFS will be aided and accelerated by the prototyping of DFS modules and VCS on the NTT. In July 1996, the NTT will be taken off the air for six months in order to install VLT-specific hardware and software for testing. From the beginning of 1997, the NTT will act like a single unit telescope. Several modules of the DFS, including the Observation Block System, will be extensively tested using SUSI as a trial instrument. All modules of the DFS will first be tested and modified within the NTT prototyping environment before going on a VLT unit telescope.

Data Pipeline Group

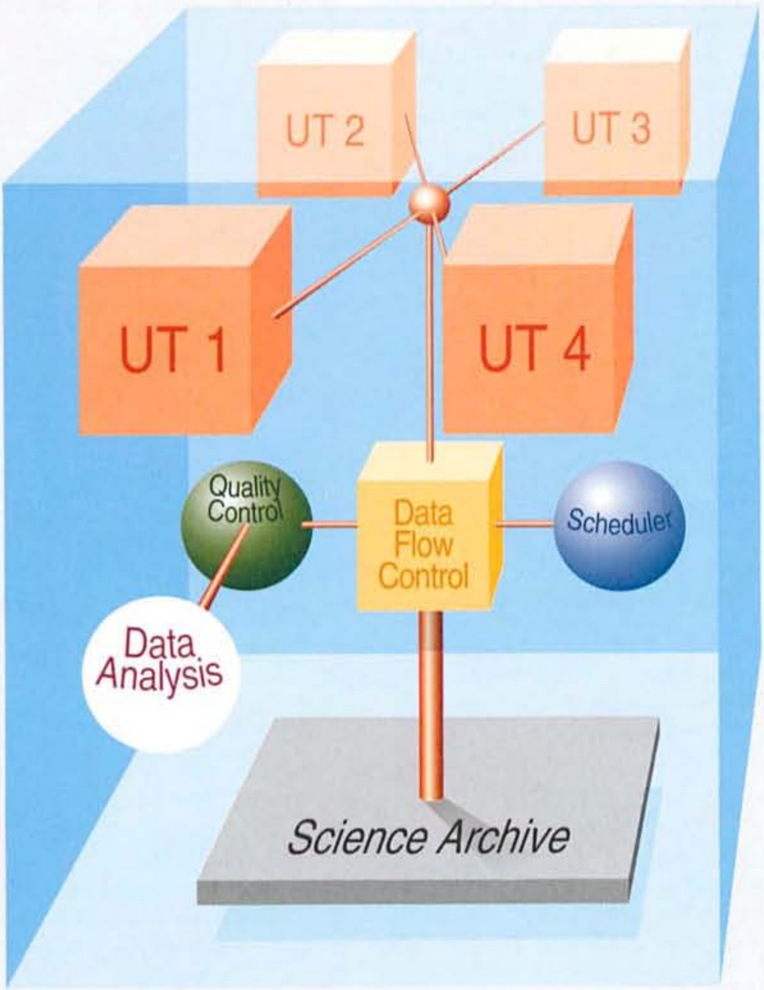
Under the leadership of Preben Grosbøl, the Data Pipeline Group is responsible for the final form of the DFS design as well as the implementation of the Pipeline and Quality Control modules of the DFS. Michèle Péron is the system designer for the DFS and is working to finalise the DFS design documents by the middle of 1996. Pascal Ballester, Klaus Banse and Carlos Guirao have responsibilities in Quality Control, Pipeline and system engineering respectively. The Data Pipeline Group is also responsible for the 95NOV release of MIDAS and the modification of MIDAS for use in the VCS. MIDAS will be used for NTT prototyping and will remain a valuable asset. Significant care has been taken to ensure that the design and high-level implementation of the Data Pipeline is independent of any specific data analysis system although MIDAS will be used for the low-level implementation on the first VLT instruments.

The VLT Science Archive

Miguel Albrecht leads the Science Archive Group which is responsible for the design and implementation of the archive. Currently the group operates the NTT archive and the European HST archive in collaboration with the Space Telescope European Co-ordinating Facility located within ESO. The VLT Science Archive will be distributed between Paranal and Garching and provide five types of services:

(a) Catalogue access for on-line and off-line needs,
(b) Storage of raw data from the VLT,
(c) Distribution of data to VLT users,
(d) Archive research on VLT and other data sets,
(e) Recalibration of raw data.

At the VLT, classical-mode astronomers and ESO staff will be able to ac-



The VLT Machine

cess catalogue information for target acquisition and observation optimisation. Raw and calibrated data will be migrated from a data cache system to Garching for long-term archive and distribution to astronomers. Cache systems on Paranal and in Garching will be of the order of 1 terabyte in capacity. The main archive in Garching will have a capacity of order 100 terabytes and consist of mixed storage media for a variety of access needs. The archive will be connected to a DFS pipeline for recalibration of raw data as required since the archive will not store calibrated and derived data in the long term. A number of other archive resources, such as the HST database, will be available to VLT archival researchers. The availability of data from ground and space-based telescopes together with a software environment of tools for the manipulation of these "meta-data" resources will provide a powerful archival research environment. The design of the VLT Science Archive will be complete in 1996 and software systems will be prototyped on the NTT in early 1997. Archive specialists Fabio Sogni and Houri Ziaeepour form the current design team.

User Support

VLT users will need support in three main areas:

(a) Proposal preparation in Phase I and II,
(b) Calibration and analysis of VLT instrument data,
(c) Archival research.

The User Support Group, based in Garching, will address these needs via services offered on the World Wide Web. The group is also responsible for general ESO WWW services and intranet facilities. As part of the observation handling system, a software team in User Support is responsible for the design and implementation of the Short, Medium and Long Term Scheduling Systems. These systems will be constructed using the SPIKE engine developed at STScI for HST scheduling. A collaborative programme between DMD and STScI will deliver a prototype of the Short Term Scheduler for testing on the NTT in early 1997. A leader for the User Support group will be identified in June this year. Currently the group consists of Maurizio Chavan, Resy de Ruijsscher and Rein Warmels.

Other DMD Activities

In January 1996, ESO entered into a contract with Serco GmbH to provide Information Technology management for the operations at Garching. Serco operates a Help Desk to assist PC and workstation users with operational problems, upgrades and maintenance. The operation of the UNIX infrastructure and networking at ESO is also Serco's responsibility. Serco provides IT planning support to ESO as well as resources for short-term projects with fixed milestones and objectives. Joe Schwarz is the ESO manager of the Outsourcing Services Centre within the DMD and acts as ESO's single point of contact to the Serco operations.

An ongoing activity now under the responsibility of DMD is the Palomar Observatory Sky Survey II plate copy project. ESO has undertaken to provide the community with film copies of the IIIa-J, IIIa-F and IV-N plates from the Palomar Schmidt telescope as part of the second-generation sky survey. Positive glass copies of the original plates are made at ESO. These plates are then used for the film production and are also sent to STScI for digital scanning. The plate taking is approximately 80% complete in IIIa-J and IIIa-F and about 40% complete for the production of film copies. Jean Quebatte leads the photographic laboratory team of Peter Dorn and Gisela Strigl.

The Next 5 Years

The testing of prototypes for the DFS on the NTT in early 1997 marks the beginning of an important five year programme for the DMD. By the end of 1997 the DFS will need to support the test camera on UT1. User Support and Science Archive resources will be in place to support both classical and service mode observing proposals on the VLT in 1998. By the end of 1998, the DFS has to be fully operational to support the first VLT instrument operations. As new VLT instruments are designed and come on-line, a prototyping and commissioning process will involve specific support for all the available instrument modes within the DFS.

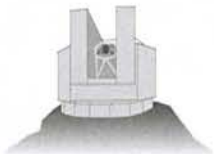
DMD will provide staff for the operation of the DFS on Paranal. Preparing and operating the scheduling system, Observation Block handling, pipeline processing, quality control of calibrations and archive operations will be carried out by DMD Data Flow Operations staff in collaboration with telescope operators, instrument scientists and staff astronomers.

The DFS is a critical component of the VLT machine that will have a fundamental bearing on the VLT's scientific productivity. The success of the DFS will depend as much on the astronomers that use it as the designers and maintainers of the system. The feedback of ideas and modifications from the user community must occur efficiently if the DFS is to provide astronomers with high-quality data products. Over the next years, DMD will host a series of workshops in which astronomers will be invited to participate in a broad range of discussions on the functionality and services offered by the DFS. These workshops will continue into the VLT operational era.

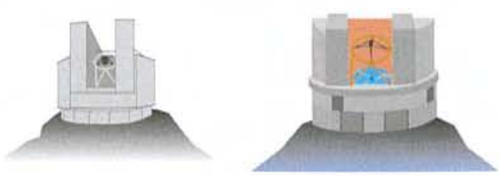
Peter Quinn
e-mail: pquinn@eso.org

DMD Major Milestones 1996–2001

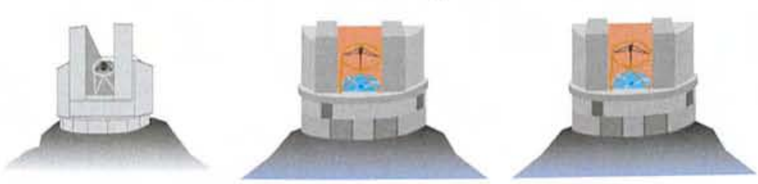
1996 Q12 • Finalise DFS design/responsibilities/interfaces
Q34 • Science Archive design document
• Construct SUSI pipeline, OB/STS & Archive prototypes for NTT



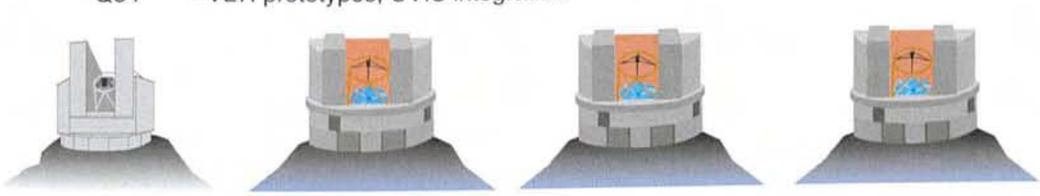
1997 Q12 • Operational prototypes on NTT, prototype development continues
• Data Flow Operations begin
Q34 • PhaseI/II operational
• Test camera supported on Paranal
• UT1 integration



1998 Q12 • DFS instrument modules operational
• STS/MTS/LTS and user support services fully operational
• Science Archive operational
Q34 • Full DFS operational
• ISAAC & FORS1 integration



1999 Q12 • CONICA integration
Q34 • VLTI prototypes, UVIS integration



2000 Q12 • VISIR & FORS 2 integration
Q34 • FUEGOS integration